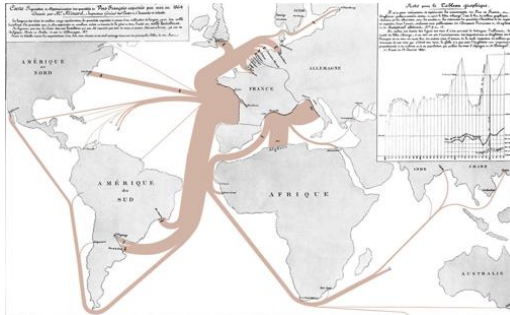


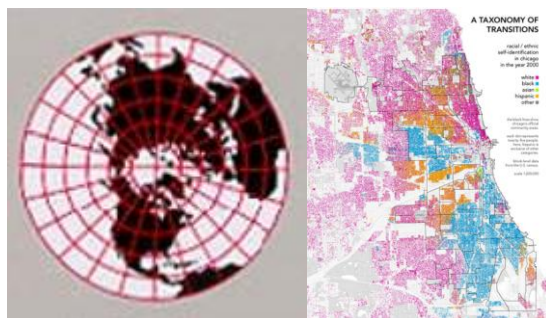
This week:

- Prepare the class for Singapore student.

1. 主要讲了空间数据和地理数据的可视化，包括时空数据的处理方法。地理位置数据针对地图的映射，点数据线数据和区域组织等进行讲解，其中的树组织算法对以后的界面布局有启发意义。递归似的从底向上组织成树结构，然后进行边绑定，可能对于以后的工程中，手机基站的轨迹绘制有很大的借鉴意义。尤其是只针对起点终点的分析。



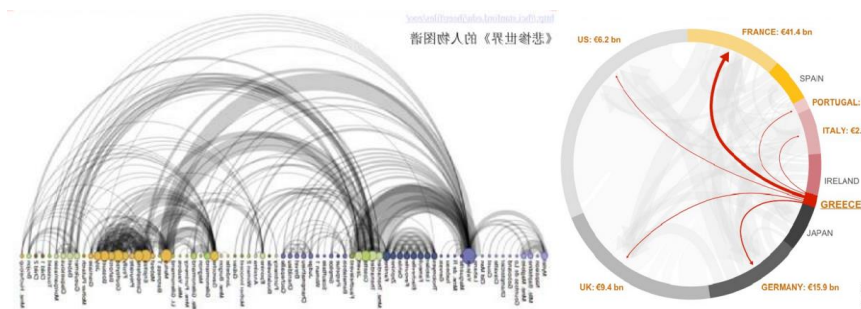
其中等距离映射方法可能可以用在显示地图上的点数据的详细视图中，采用等距离的辅助线在绘制详细地图，新的工作中可以尝试一下。



关于点的重叠问题，可能可以借鉴一下 **PixelMap** 算法，但是我们的可视化中可能强调点的位置精确，所以肯能这个算法可以对基站数据的 **voronoi** 图进行手机轨迹数据的可视化，这样手机轨迹的点就不是都集中在一个位置上了。效果可能做出来如上图。

2. 其次还有关系数据的可视分析和设计。主要还是针对树结构进行讲解，介绍了旭日图等并分析了优劣性。其中正交图视觉效果比较好，但是算法复杂。环形排列布局比较适合展示个人的社交关系，但是当叶节点数目增多时就需要一些策略来防止叶节点过小。对于层次结构强的还是介绍了嵌套的排列。还介绍了矩阵图。

关系数据来说，个人认为弧形视图不如圆环。显得比较乱。空间利用也不够

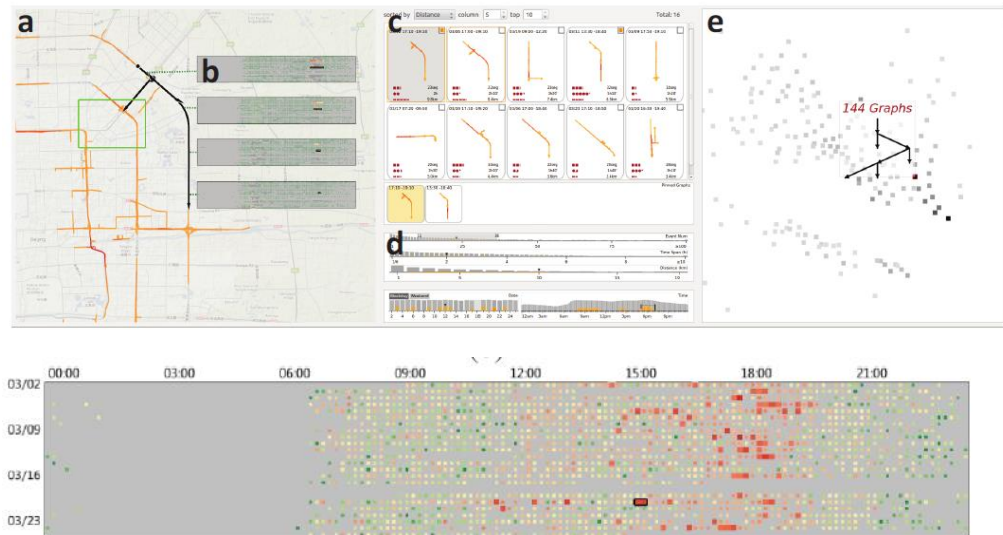


这一部分对手机社交关系数据可能有借鉴意义。

● Read 5 paper.

1. Visual Traffic Jam Analysis Based on Trajectory Data

本文主要想要解决的问题是交通拥堵问题的可视分析。



首先计算了每一条道路上的出租车的速度，然后找到速度较低的道路作为分析对象。首先在地图上用颜色编码拥堵，并把热力图映射到路网上。相信信息上，采用了矩阵图来表示这些计算出的速度。每十分钟一个时间段。可以看出上图中的道路有明显的晚高峰现象。最有亮点的工作出现在本文利用路网，研究了拥堵的传播关系。虽然方法并不是自动的。。。其他的可视化编码都比较简单，但是颜色选取还是让人感觉很舒服的。

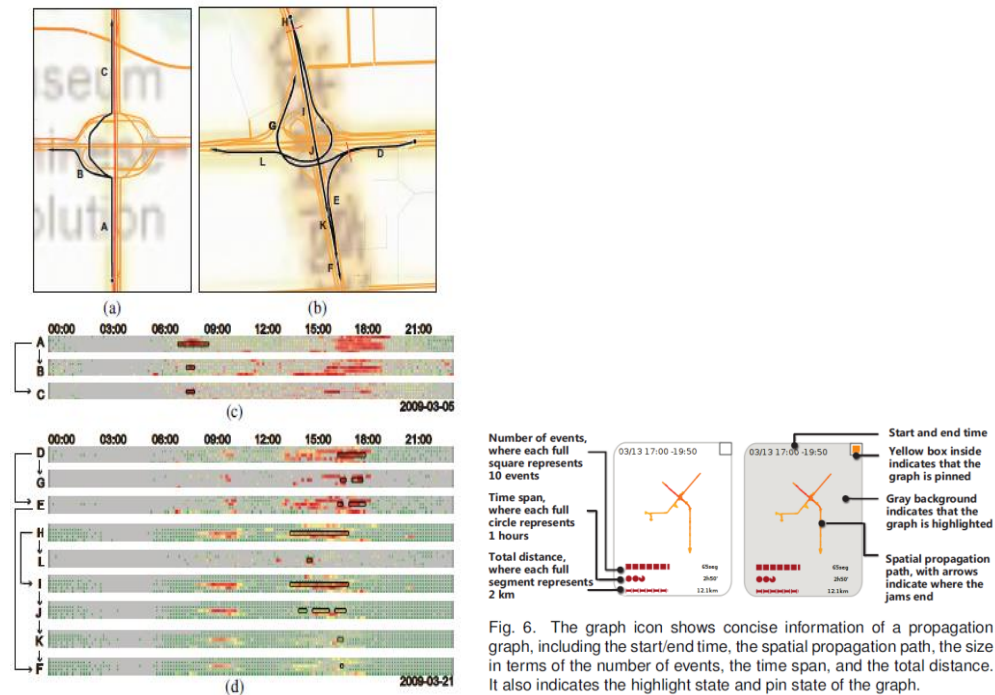


Fig. 6. The graph icon shows concise information of a propagation graph, including the start/end time, the spatial propagation path, the size in terms of the number of events, the time span, and the total distance. It also indicates the highlight state and pin state of the graph.

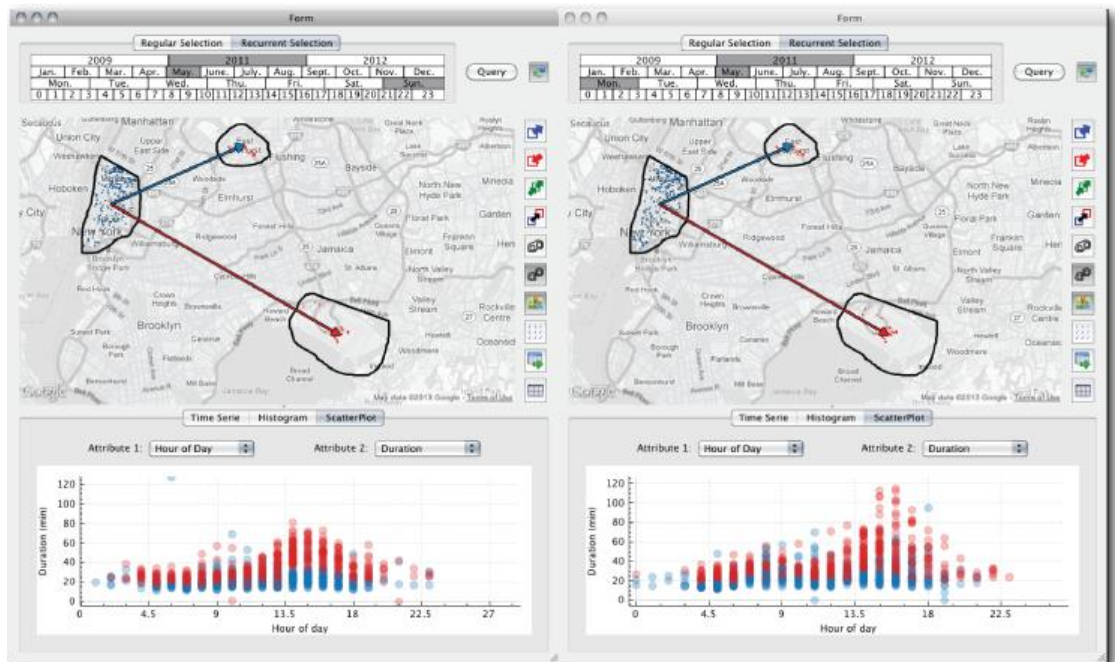
2. Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York

City Taxi Trips

这篇文章之前大致的看过，本文想要解决的难题主要是有两个：第一是查询，查询现有的工具大都是基于数据库语言的，没有易懂的可视化方法。第二是针对数据量，因为数据量还是很大的，平常的工作都是使用小的数据样本来分析。

本文也是考虑了用数据库的查询方法，建立索引。但是时间周期太长，所以和我们一样建立了轻量级的数据库，基于 KDtree 来加速查询，并缩小存储空间。

采用的查询系统基本就是 VAUD 的一个条件节点来建立时间空间和属性的三元关系查询，支持对于查询结果的统计可视化。



首先系统主要提供 **overview** 的可视分析任务，主要是针对出租车数量和上下客地点：



以及系统对于区域的模式分析，还是针对上下客地点车数量的周期性的模式分析。

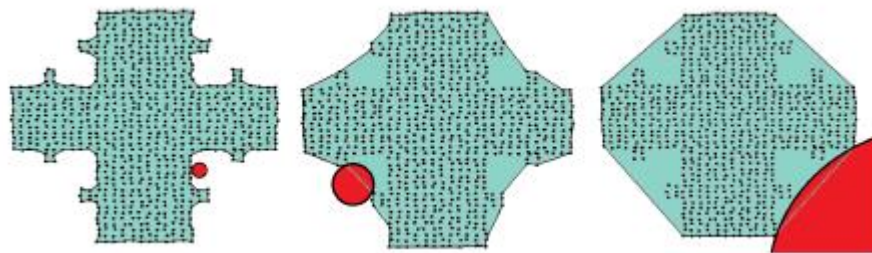


此外可以查看某区域上车的出租车的下车地点分布。总体来说是一个针对 OD 轨迹数据的可视化方法。其中的多轨迹的统计可视化可以借鉴。

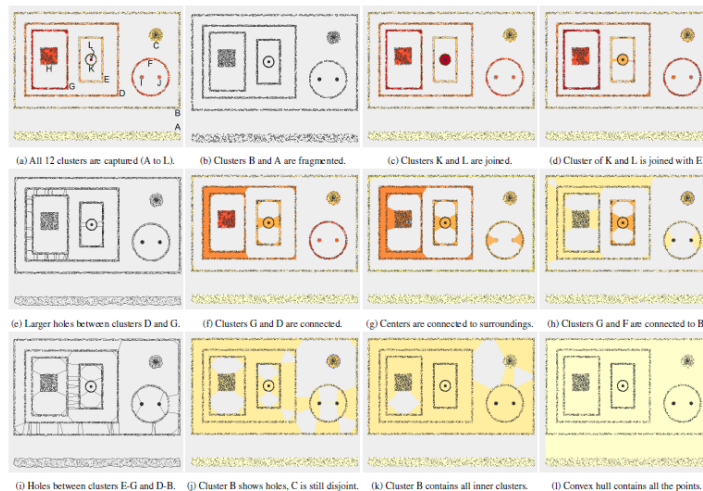
3. Visual Analytics for Spatial Clustering: Using a Heuristic Approach for Guided Exploration

这篇文章主要针对地理数据的空间聚类进行可视化，主要是想要能够动态的调整聚类参数，并且能够展示聚类结果。系统采用启发式的方法完成对空间聚类的探索。

本文采用的聚类方法 α -sharp 主要思想是用半径为 α 的圆来铺在平面中，园中不能有数据点，然后边界随着圆的数量增多而逐渐形成约束，达到聚类效果，如下图是三个不同值的 α 的聚类过程。



把聚类后的不同的不连通的区域用不同的颜色表示，完成对数据空间的聚类。K 阶的算法表示圆可以与数据点有 k 个交点。

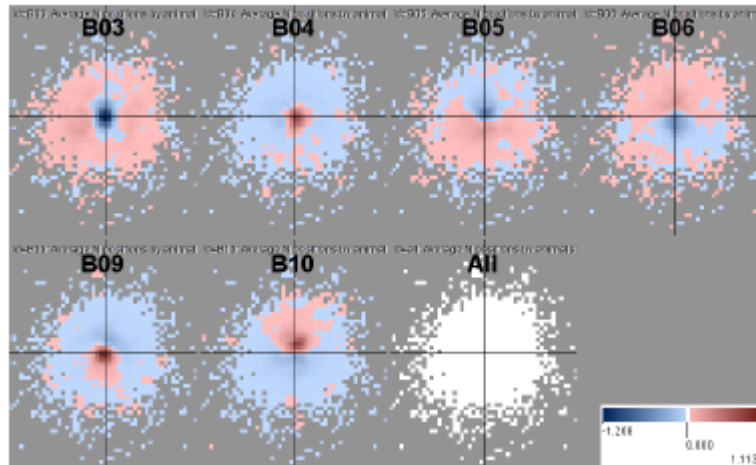


上图中还有没有颜色的图，没有颜色的图只保留了聚类后的边界，而如果有颜色则表示聚类中没有单个数据点的聚类并且聚类的边界是简单的。本文还采用热力图来表示时间维度，表示随着 α 值得变化，聚类的变化情况。

本文的聚类方法可能可以用于对于行政区域或者人群聚集地的划分聚类。

4. Space Transformation for Understanding Group Movement

这篇那篇猩猩移动的文章，精度之后发现文章主要是想要研究移动数据的移动模式，并专注于起领导作用的个体。采用 STC 模型，可能是考虑之后用于大数据。主要的方法是把每一个个体的运动拆分成平均运动轨迹和相对于平均轨迹的运动和。进而计算分析群体中的个体行为：



6 只猩猩的比较图，发现第一只乱走，第二只主要在群体中间。

本篇文章的分析可以运用在手机轨迹数据聚类后对于单个聚类的详细分析。里面有有意思的聚类内分析方法，可能可以在朱博的聚类工作上。思路比较新颖，聚类内单体的分析，是不是能够在聚类内再聚类分析（为了隐私保护）

5. AllAboard Visual Exploration of Cellphone Mobility Data to Optimise Public Transport

本文的工作和大师姐的投稿有点像。文章框架为底层路网信息和轨迹数据被对象化（人），然后进行轨迹分析或轨迹聚类。包括 OD 检测，停留点监测，公共路径等，然后提供给前台 API 来进行过滤，优化，分型等操作。

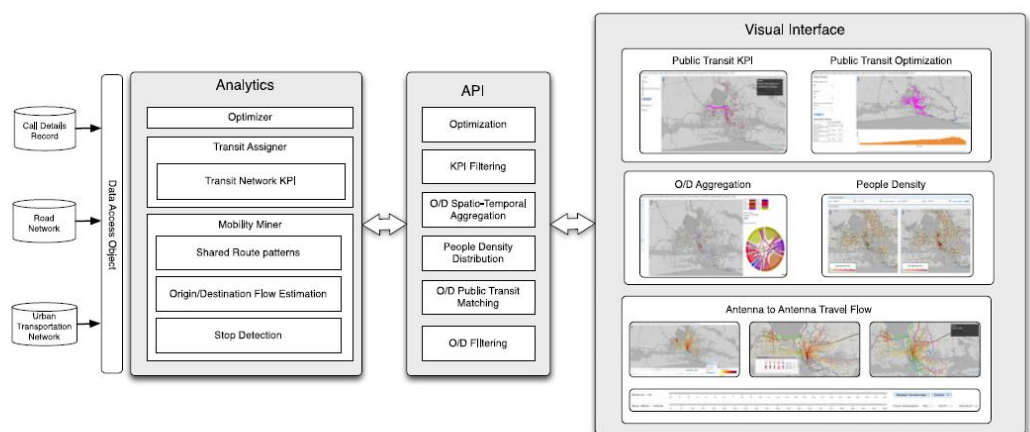


Fig. 1. Architecture of the AllAboard platform.

● 思路

1. 现在的查询系统可能是针对全局的数据展示，而且是针对用户已经大致

知道需要查询什么样的结果。

探索城市数据，数据量比较大。我想根据用户关注的，利用所见即世界的唯心主义（巫老师的同学报告过的心理学的文章），做一个以鼠标为视点的可视化工具，就好像鼠标是一个人，他在已有的时空数据中探索，就像一个时空漫游者（因为大多数时间用户的视点跟着鼠标的）。实时的展示数据。并根据鼠标的点击来进行数据的过滤。编程的困难性就在于如何做到实时性。

有一个思路就是当数据量大的时候，展示所有的数据在地图上也会出现遮挡，不如进行采样或统计的方法。这个还是要搭建好数据库后进行全局的编程规划。这次数据处理一定要全程放在后台完成。

2. 第二个想法是探测异常数据。出租数据有上下车，当乘客在车上时是否有绕路异常，此类方法可以延伸到轨迹的异常处理。
3. 交通流量预测。比较偏重机器学习，把数据作为训练集来预测交通的流量。或者规划城市道路。

- Translate the paper.

翻译了一篇关于机器学习中的空间转换模块的文章，他把神经网络中加入了一个空间转换器模块使得神经网络可以对图片进行空间变化。这和我在投稿月的时候有一个想法比较类似，就是把程序模块功能化，虽然可能不能发表文章，但是可以在以后的编程设计中实现，不浪费时间，并且以后的程序也能直接使用封装好的 js 功能模块，缩短程序实现效率。同时对于新加入实验室的本科同学，也可以写用几个封装好的模块写的小工具，然后让他们读代码，快速入门。

- Study algorithm and programming with python.

1. 曲线拟合：数据采样和采样点曲线拟合
2. PCA：降维投影，这个实现起来比较简单

Next week:

- Go to class 23.
- Read more 5 paper about Traj.
- Research Internet of things.